

PENGARUH PEMBOBOTAN EMOJI TERHADAP EVALUASI ALGORITME NAÏVE BAYES PADA KOMENTAR PELECEHAN SEKSUAL

Ismi Damayanti, Tesa Nur Padilah, Ultach Enri

Program Studi Informatika S1, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang
Jl. HS. Ronggo Waluyo, Telukjambe Timur, Karawang, Indonesia
1910631170088@student.unsika.ac.id

ABSTRAK

Sebuah akun dengan username @kuydotid mengunggah video tentang berita kasus pelecehan seksual Universitas Gunadarma. Dengan munculnya banyak komentar dari pengguna Tiktok tentang kejadian tersebut, maka pendapat-pendapat tersebut perlu dianalisis untuk mengklasifikasikan sentimen yang terkandung di dalamnya. Penelitian ini menambahkan pembobotan opini non-tekstual berupa emoji untuk menguji pengaruh terhadap hasil akurasi. Penambahan pembobotan emoji ini dilakukan guna menghindari makna ambigu pada sebuah komentar. Untuk penelitian ini, objek yang akan diteliti adalah komentar pengguna dari sosial media Tiktok pada unggahan @kuydotid tentang kasus pelecehan seksual di Universitas Gunadarma. Data tersebut berisi teks dan emoji (non-tekstual) untuk dianalisis sentimen. Pada penelitian ini, metodologi yang digunakan adalah *Knowledge Discovery in Database* (KDD). Pada penelitian ini evaluasi data mining menggunakan *confusion matrix* yang menghasilkan nilai *accuracy*, *precision* dan *recall*. Penggunaan pembobotan emoji telah terbukti memiliki pengaruh positif terhadap pengklasifikasian dan akurasi sistem. Hal ini dapat dilihat dari peningkatan nilai evaluasi.

Kata kunci: Klasifikasi, Pelecehan Seksual, Pembobotan Emoji

1. PENDAHULUAN

Perkembangan teknologi telah terjadi begitu cepat dan memberikan banyak manfaat, salah satunya adalah informasi lebih mudah dan cepat. Sebanding dengan perkembangan yang cepat dalam teknologi, masyarakat dalam lingkungan sosial telah menguasai ilmu dan teknologi penggunaan internet, khususnya media sosial [1].

catatan laporan survei *Reuters Institute Digital News Report* (2022) menemukan data sebanyak 16% dari responden di Indonesia mengandalkan Tiktok sebagai sumber informasi. Menurut laporan *We Are Social dan Hootsuite* pada bulan Januari 2023, Tiktok telah mencapai jumlah pengguna sebesar 1,05 miliar di seluruh dunia. Pada bulan Januari, Tiktok berada di peringkat keenam sebagai aplikasi media sosial dengan jumlah pengguna terbanyak di seluruh dunia. Jumlah pengguna Tiktok secara global meningkat sebesar 18,8% dibandingkan dengan tahun sebelumnya.

Tiktok merupakan salah satu aplikasi berita berbasis android yang menjadi tempat banyak orang mencari dan menyebarkan berita terbaru. Terdapat banyak pilihan berita yang dapat diakses dengan mudah, mulai dari berita di Indonesia hingga berita mancanegara. Salah satu topik berita yang sedang menarik perhatian publik saat ini adalah tentang kasus pelecehan seksual, termasuk kasus pelecehan seksual di commuter line stasiun Sudimara dan kekerasan seksual yang diterima 13 santriwati di Bandung. Sebagaimana dijelaskan kekerasan seksual diartikan sebagai setiap perilaku yang tertuju pada seksualitas atau organ seksual seseorang tanpa persetujuan mereka, dengan unsur paksaan atau ancaman,

termasuk pelecehan seksual dan perdagangan perempuan untuk tujuan seksual [2].

Hasil survei *International NGO Forum on Indonesian Development* (2021) menemukan bahwa 71,8 persen masyarakat Indonesia pernah mengalami kekerasan seksual baik terhadap diri sendiri, keluarga maupun orang lain yang mereka kenal. Dari jumlah tersebut, 65,1% dari mereka yang disurvei menyatakan pernah mengalami pelecehan seksual, ini merupakan jenis tertinggi yang sering kita kenali. Upaya untuk memerangi pelecehan seksual masih terus diusung oleh masyarakat hingga saat ini. Data yang tersedia menunjukkan bahwa terdapat 338.496 kasus kekerasan seksual yang tercatat [3].

Universitas menduduki urutan paling tinggi dalam kasus pelecehan seksual diantara jenjang pendidikan lain. Laporan oleh Komisi Nasional Anti Kekerasan Terhadap Perempuan menunjukkan bahwa universitas adalah lembaga yang paling banyak menerima pengaduan kekerasan seksual. Terhitung sampai 2022 terdapat 173 kasus pelecehan seksual di ranah publik yang berasal dari jenjang pendidikan di universitas [3].

Pada akhir bulan Desember 2022, terjadi kasus pelecehan seksual terhadap salah satu mahasiswa di Universitas Gunadarma Depok yang ramai dibahas di Tiktok. Sebuah akun dengan username @kuydotid mengunggah video tentang berita kasus pelecehan seksual Universitas Gunadarma. Dengan munculnya banyak komentar dari pengguna Tiktok tentang kejadian tersebut, maka pendapat-pendapat tersebut perlu dianalisis untuk mengklasifikasikan sentimen yang terkandung di dalamnya. Untuk mengelompokkan unggahan dari pengguna ke dalam

kelas positif dan negatif secara otomatis, metode yang digunakan disebut analisis sentimen.

Berdasarkan permasalahan tersebut maka pada penelitian ini akan dilakukan penerapan metode klasifikasi *Naive Bayes* untuk menentukan klasifikasi sentimen positif dan negatif pada unggahan Tiktok tentang pelecehan seksual. Untuk menghindari makna ambigu pada komentar atau ketidaksesuaian antara komentar dan emoji, maka penelitian ini memerlukan penambahan pembobotan opini non-tekstual berupa emoji untuk menguji pengaruh terhadap hasil evaluasi

2. TINJAUAN PUSTAKA

2.1. Tiktok

Tiktok adalah sebuah aplikasi yang menawarkan efek khusus yang unik dan menarik, memungkinkan pengguna dengan mudah membuat video pendek yang menarik dan dapat menarik perhatian banyak orang yang melihatnya. Video pendek yang dibuat dapat dibagikan kepada teman-teman di media sosial dan pengguna Tiktok lainnya [1].

2.2. Pelecehan Seksual

Pelecehan seksual adalah suatu bentuk kekerasan yang dilakukan dengan cara menggunakan kekuatan fisik, kekuasaan, atau ancaman terhadap orang lain dengan tujuan untuk memuaskan hasrat seksual seseorang [2].

2.3. Data Mining

Data mining adalah proses mengidentifikasi pola dari data yang besar menggunakan teknik-teknik pembelajaran mesin dan analisis statistik [4]. Tujuan dari *data mining* adalah untuk menemukan informasi yang berguna dari data yang besar, dengan cara yang efisien dan otomatis. *Data mining* digunakan untuk mengekstrak fitur-fitur yang penting dari sebuah dataset menggunakan metode seleksi fitur [5].

2.4. Text Mining

Text mining adalah suatu proses yang digunakan untuk mengekstrak informasi dari teks yang besar menggunakan teknik-teknik *text mining*. *Text mining* sering digunakan untuk mengekstrak informasi struktural maupun tidak struktural dari teks, seperti entitas nama, hubungan antar entitas, atau tema yang terkandung dalam teks tersebut [6].

2.5. Analisis Sentimen

Analisis sentimen adalah proses untuk mengekstrak informasi mengenai perasaan atau opini yang terkandung dalam suatu teks. Analisis sentimen sering digunakan untuk mengukur reaksi atau tanggapan terhadap produk, layanan, atau peristiwa yang terdapat dalam teks [7].

2.6. Python

Python adalah sebuah bahasa pemrograman yang dikembangkan oleh Guido van Rossum dan dikenal sebagai bahasa pemrograman skrip dan pemrograman

web yang populer. *Python* adalah bahasa pemrograman interpretatif yang serbaguna dengan filosofi desain yang berfokus pada tingkat keterbacaan kode. *Python* dianggap sebagai bahasa yang menggabungkan kemampuan dan fungsionalitas yang kuat dengan sintaksis kode yang sangat jelas. Selain itu, *Python* dilengkapi dengan pustaka standar yang luas dan komprehensif [8].

2.7. Naïve Bayes Classifier

Naive Bayes adalah sebuah metode klasifikasi yang berdasarkan teori probabilitas dan statistik. Metode ini mengasumsikan bahwa setiap fitur yang ada dalam data bersifat independen [9].

2.8. Pembobotan Emoji

Emoji adalah simbol grafis unicode yang digunakan untuk mengekspresikan sentimen tertentu dalam dokumen. Oleh karena itu, pembobotan *emoji* sangat penting untuk menentukan sentimen positif atau negatif secara jelas. Pembobotan emoji dilakukan bersamaan dengan pengklasifikasian tekstual. Data teks pertama-tama diidentifikasi untuk mengetahui apakah mengandung *emoji* atau tidak berdasarkan kamus *emoji* yang disimpan dalam database. Kemudian, semua *emoji* yang terkandung dalam sentimen dijumlahkan untuk memberikan bobot pada *emoji* yang bersangkutan, baik positif maupun negatif [10].

2.9. Knowledge Discovery in Databases

Knowledge Discovery in Databases (KDD) adalah proses pengenalan pola dari data yang besar dan kompleks dengan tujuan untuk menemukan informasi berharga yang dapat digunakan untuk memecahkan masalah [11].

2.10. Transformation TF-IDF

Transformasi TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah metode yang digunakan dalam pemrosesan teks untuk memberikan bobot pada kata-kata dalam dokumen berdasarkan frekuensi kemunculannya dalam dokumen tersebut dan seluruh korpus teks yang ada [12]. Pada tahap transformasi TF-IDF, data teks dibagi menjadi dua bagian: training data dan testing data. Selanjutnya, dilakukan perhitungan bobot kata dengan menggunakan metode TF-IDF pada data training.

2.11. Confusion Matrix

Confusion matrix dijelaskan sebagai sebuah tabel yang digunakan untuk menggambarkan performa suatu model klasifikasi [13]. *Confusion matrix* dapat digunakan untuk menghitung beberapa metrik evaluasi model klasifikasi, yaitu *accuracy*, *precision*, dan *recall*. *Accuracy* adalah sebuah metrik evaluasi model klasifikasi yang menggambarkan seberapa baik model tersebut dapat memprediksi kelas dari suatu data. *Precision* menggambarkan seberapa baik model dapat memprediksi kelas positif. Sedangkan *recall*

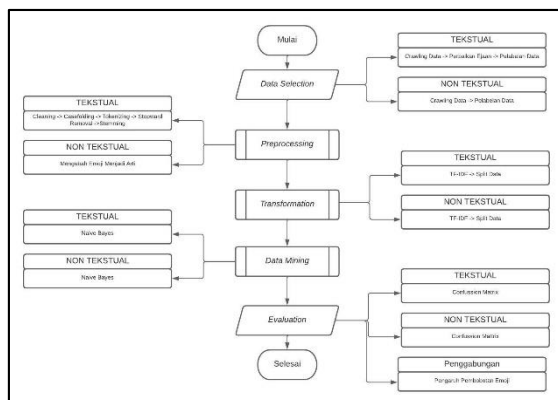
menggambarkan seberapa baik model dapat mengidentifikasi semua kelas positif yang sebenarnya ada).

2.12. Penggabungan Pembobotan

Keputusan akhir klasifikasi untuk menentukan sentimen dalam analisis sentimen yang menggabungkan pembobotan tekstual dan non-tekstual diperoleh dari penggabungan kedua pembobotan tersebut [14].

3. METODE PENELITIAN

Metodologi penelitian merupakan cara untuk mengetahui langkah-langkah yang akan diambil untuk menghasilkan informasi yang sesuai dengan tujuan penelitian. Pada penelitian ini, metodologi yang digunakan adalah *Knowledge Discovery in Database* (KDD).



Gambar 1. Rancangan Penelitian

4. HASIL DAN PEMBAHASAN

Pada tahap *data selection* dimulai dengan mengumpulkan informasi berupa komentar pengguna pada unggahan @kuydotid menggunakan alat *Instant Data Scraper*. Pengambilan data diambil secara satu kali pada rentang waktu komentar 13 Desember-20 Desember 2022. Data yang berhasil dikumpulkan sebanyak 383 komentar. Setelah melakukan pengumpulan data dari komentar pada unggahan @kuydotid, terdeteksi beberapa kesalahan penulisan bahasa yang tidak sesuai dengan standar dalam Kamus Besar Bahasa Indonesia (KBBI). Untuk perbaikan kesalahan ejaan tersebut, divalidasi oleh seorang ahli tata bahasa yang merupakan lulusan sarjana Pendidikan Bahasa Sastra Indonesia.

Dalam upaya membantu proses klasifikasi, diperlukan adanya label atau kategori yang dapat membantu algoritme mengenali kelas-kelas tertentu. Namun, karena dataset yang digunakan masih dalam bentuk mentah dan tidak memiliki label, maka dilakukan pelabelan secara manual pada setiap komentar pada dataset. Pada tahap ini, seorang ahli tata bahasa yang merupakan lulusan sarjana Pendidikan Bahasa dan Sastra Indonesia melakukan validasi label

pada setiap dataset yang telah melalui tahapan sebelumnya.

Tabel 1. Ringkasan validasi label sentimen

Sebelum Validasi Ahli		Setelah Validasi Ahli	
Komentar	Sentimen	Komentar	Sentimen
Waduh kan viral	NEGATIF	Waduh kan viral	POSITIF
Depok lagi	NEGATIF	Depok lagi	NETRAL
Kasian anjir, mana baru banget ulangan tengah semester. malah di seret ke pohon	NEGATIF	Kasian anjir, mana baru banget ulangan tengah semester. malah di seret ke pohon	POSITIF

Dari total 383 data yang ditemukan dan telah dilakukan validasi oleh ahli, terdapat 105 ulasan dengan label positif, 220 ulasan dengan label netral dan 58 ulasan dengan label negatif. Dari 383 komentar terdapat 46 komentar yang mengandung emoji. Label pada komentar tersebut terdiri dari 18 komentar positif 20 komentar netral dan 8 komentar negatif.

Selanjutnya, pada seleksi data non-tekstual, dari 383 komentar Tiktok pada akun @kuydotid terdapat 46 komentar yang menggunakan emoji. Pada dataset tersebut dilakukan penghapusan berupa komentar tekstual dan hanya menggunakan komentar non-tekstual saja berupa emoji. Dari total 46 data yang ditemukan, terdapat 30 ulasan dengan label positif, 6 ulasan dengan label netral dan 10 ulasan dengan label negatif.

Setelah *data selection*, tahapan selanjutnya yaitu *preprocessing*. Pada tahapan pertama *preprocessing* yaitu cleaning dan casefolding dilakukan penghapusan beberapa karakter dan konversi semua kata menjadi huruf kecil. Langkah selanjutnya adalah tokenisasi. Pada tahap ini, kata-kata akan dipisahkan satu per satu. Lalu pada tahap *stopword removal* dilakukan penghapusan kata-kata umum yang tidak memiliki arti. Setelah itu data akan melewati tahap stemming dengan mengembalikan kata-kata ke bentuk dasarnya dengan menghilangkan imbuhan. Sedangkan pada *preprocessing non-tekstual* emoji diartikan ke dalam string dengan validasi ahli.

Tahap selanjutnya yaitu transformasi. Pada transformasi tekstual dilakukan pembagian data menjadi training data dan testing data dan metode pembobotan yang digunakan yaitu TF-IDF (Term Frequency-Inverse Document Frequency). Setelah data berubah menjadi bentuk yang dipahami algoritme, dilakukan split data dengan rasio pembagian data sebesar 80:20 dan 70:30. Sedangkan pada transformasi data non-tekstual emoji yang sudah divalidasi menjadi bentruk string dilakukan pembobotan.

Pada tahapan data mining dilakukan klasifikasi sentimen terhadap komentar pada unggahan @kuydotid di Tiktok tentang kasus pelecehan seksual universitas menggunakan algoritme *Naïve Bayes*.

accuracy:	0.500				
	precision	recall	f1-score	support	
negatif	0.00	0.00	0.00	2	
netral	0.40	0.67	0.50	3	
positif	0.67	0.67	0.67	3	
accuracy			0.50	8	
macro avg	0.36	0.44	0.39	8	
weighted avg	0.40	0.50	0.44	8	

Gambar 2. Hasil *data mining* tekstual 80:20

Hasil data mining tersebut merupakan 20% data dari 46 komentar dengan rasio perbandingan *split* data 80:20. Data testing tersebut menghasilkan sentimen positif sebanyak 3 komentar dengan presentase 37,5%, sentimen netral sebanyak 3 komentar dengan presentase sebesar 37,5% dan ulasan sentimen negatif sebanyak 2 komentar dengan presentase sebesar 25%.

accuracy:	0.417				
	precision	recall	f1-score	support	
negatif	0.00	0.00	0.00	2	
netral	1.00	0.29	0.44	7	
positif	0.30	1.00	0.46	3	
accuracy			0.42	12	
macro avg	0.43	0.43	0.30	12	
weighted avg	0.66	0.42	0.37	12	

Gambar 3. Hasil *data mining* tekstual 70:30

Data testing pada rasio *split* data 70:30 tersebut menghasilkan sentimen positif sebanyak 2 komentar dengan presentase 16,67%, sentimen netral sebanyak 7 komentar dengan presentase sebesar 58,33% dan ulasan sentimen negatif sebanyak 3 komentar dengan presentase sebesar 25%. Kemudian, pada data non-tekstual juga dilakukan *split* data 80:20 dan 70:30.

accuracy:	0.900				
	precision	recall	f1-score	support	
negatif	1.00	1.00	1.00	1	
netral	0.00	0.00	0.00	1	
positif	0.89	1.00	0.94	8	
accuracy			0.90	10	
macro avg	0.63	0.67	0.65	10	
weighted avg	0.81	0.90	0.85	10	

Gambar 4. Hasil *data mining* non-tekstual 70:30

Hasil tersebut diperoleh melalui 20% *data testing* dari 46 komentar yaitu sebanyak 10 komentar. Data mining menghasilkan komentar sentimen positif sebesar 80% atau sebanyak 8 komentar, sentimen netral sebesar 10% sebanyak 1 komentar dan komentar sentimen negatif sebesar 10% atau setara dengan 1 komentar.

accuracy:	0.929				
	precision	recall	f1-score	support	
negatif	1.00	1.00	1.00	2	
netral	0.00	0.00	0.00	1	
positif	0.92	1.00	0.96	11	
accuracy			0.93	14	
macro avg	0.64	0.67	0.65	14	
weighted avg	0.86	0.93	0.89	14	

Gambar 5. Hasil *data mining* non-tekstual 70:30

Data testing pada rasio *split* data 70:30 tersebut menghasilkan sentimen positif sebanyak 2 komentar dengan presentase 14,29%, sentimen netral sebanyak 1 komentar dengan presentase sebesar 7,14% dan ulasan sentimen negatif sebanyak 3 komentar dengan presentase sebesar 78,57%.

Pada tahapan terakhir yaitu evaluasi terdapat 3 proses pada tahap ini, yaitu evaluasi untuk sentimen tekstual, non-tekstual dan penggabungan evaluasi. Tahap evaluasi adalah tahap dimana performa implementasi algoritme *Naïve Bayes Classifier* dievaluasi menggunakan *confusion matrix*. Hasil dari perbandingan rasio *split* data pada penelitian ini menunjukan bahwa 80:20 menghasilkan nilai evaluasi yang lebih tinggi daripada rasio 70:30.

Accuracy Score: 50.00
Precision Score: 35.56
Recall Score: 44.44

Gambar 6. Hasil evaluasi tekstual 80:20

Accuracy Score: 41.67
Precision Score: 43.33
Recall Score: 42.86

Gambar 7. Hasil evaluasi tekstual 70:30

Hasil evaluasi komentar berupa *emoji* dengan menggunakan *confusion matrix* dan menghasilkan kesimpulan bahwa Hasil dari perbandingan rasio *split* data pada penelitian ini menunjukan bahwa 70:30 menghasilkan nilai evaluasi yang lebih tinggi daripada rasio 80:20.

Accuracy Score: 90.00
Precision Score: 62.96
Recall Score: 66.67

Gambar 8. Hasil evaluasi non-tekstual 80:20

Accuracy Score: 92.86
Precision Score: 63.89
Recall Score: 66.67

Gambar 9. Hasil evaluasi non-tekstual 70:30

Langkah selanjutnya yaitu dilakukan penggabungan evaluasi untuk mengetahui apakah emoji memiliki pengaruh signifikan terhadap hasil evaluasi sistem secara keseluruhan. Dengan melakukan pengujian ini, diharapkan juga dapat ditemukan komposisi konstanta perbandingan nilai pembobotan teks dan non-teks yang optimal.

Tabel 2. Pengujian pengaruh pembobotan emoji

Konstanta		Accuracy	Precision	Recall
A	B			
1	0	50%	35,56%	44,44%
0.9	0.1	54,29%	38,21%	46,66%
0.8	0.2	58,57%	41,07%	48,89%
0.7	0.3	62,86%	43,92%	51,11%
0.6	0.4	67,14%	46,77%	53,33%
0.5	0.5	71,43%	49,63%	55,56%
0.4	0.6	75,72%	52,48%	57,78%
0.3	0.7	80%	55,33%	60%
0.2	0.8	84,29%	58,18%	62,22%
0.1	0.9	88,57%	61,04%	64,45%
0	1	92,86%	63,89%	66,67%

Berikut adalah hasil pengujian akurasi sistem dengan menggunakan pembobotan tekstual saja, pembobotan non-tekstual saja, dan penggabungan pembobotan tekstual dan non-tekstual.

Tabel 3. Penggabungan Evaluasi Pengaruh Pembobotan Emoji

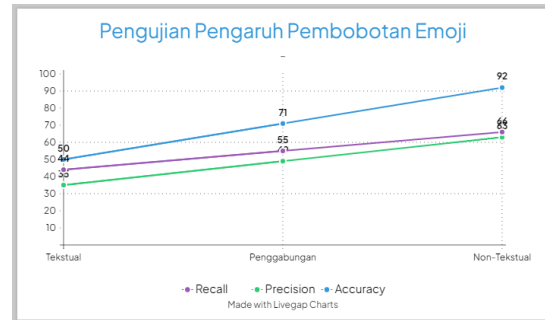
Evaluasi	Tekstual (α)	Non-Tekstual (β)	Penggabungan (Konstanta $\alpha = 0,9$ dan $\beta = 0,1$)
Accuracy	50%	92,86%	71,43%
Precision	35,36%	63,89%	49,63%
Recall	44,44%	66,67%	55,56%

Pada kondisi pengujian dengan menggunakan pembobotan tekstual saja, hasil akurasi yang diperoleh adalah sebesar 50%. Setelah melakukan penggabungan pembobotan emoji hasil akurasi meningkat menjadi 71,43%.

Hal yang sama terjadi pada hasil evaluasi lain seperti nilai *precision* dan *recall*. Pada nilai *precision* tekstual senilai 35,36% terjadi kenaikan menjadi 49,63% setelah dilakukan penggabungan pembobotan emoji. Begitupun nilai *recall*, untuk nilai tekstual sebesar 44,44% dan terjadi lonjakan pada penggabungan nilai *recall* sebesar 55,56%.

Penggunaan pembobotan emoji telah terbukti memiliki pengaruh positif terhadap pengklasifikasian dan akurasi sistem.

Pada gambar 10 menunjukkan grafik yang meningkat. Dengan demikian, dapat dikatakan bahwa pembobotan non-tekstual memiliki pengaruh signifikan terhadap hasil klasifikasi sentimen.



Gambar 10. Pengujian pengaruh pembobotan emoji

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dapat diperoleh kesimpulan bahwa klasifikasi sentimen pada komentar Tiktok di akun @kuydotid tentang kasus pelecehan seksual dengan 383 komentar tekstual, menghasilkan 105 komentar positif dengan presentase 27,42%, 220 komentar netral dengan presentase 57,44% dan 58 komentar negatif dengan presentase 15,14%. Pada penelitian ini juga terdapat hasil klasifikasi sentimen komentar non-tekstual dengan jumlah dataset sebanyak 46 komentar, terdiri dari 30 komentar non-tekstual positif dengan presentase 65,22%, 6 sentimen netral senilai dengan 13,04% dan 10 komentar non-tekstual negatif dengan presentase 21,74%. Performa algoritme *Naïve Bayes Classifier* dengan penambahan pembobotan emoji memberikan pengaruh baik terbukti dengan kenaikan persentase nilai *accuracy*, *precision*, dan *recall*, adapun saran untuk pengembangan penelitian yaitu diharapkan menggunakan dataset yang memiliki banyak sentimen non-tekstual (emoji) supaya tidak terjadi penurunan data signifikan dari data awal pada saat proses *crawling* dan mengeksplorasi rasio *split* dan metode evaluasi yang berbeda.

DAFTAR PUSTAKA

- [1] Aprilian, D., Elita, Y., & Afriyati, V. (2019). Hubungan antara Penggunaan Aplikasi Tiktok dengan Perilaku Narsisme pada Siswa Sekolah Menengah Pertama. *Consilia Jurnal Ilmiah BK*, 2(3).
- [2] World Health Organization, 2022. <https://www.who.int/news-room>. (accessed Feb. 23, 2022).
- [3] UNICEF, "Every child is protected from violence and exploitation," *Glob. Annu. Results Rep. 2021*, pp. 1–137, 2021.
- [4] Devita, R. N., Herwanto, H. W., & Wibawa, A. P. (2018). Perbandingan kinerja metode naive bayes dan k-nearest neighbor untuk klasifikasi artikel berbahasa indonesia. *JTIK (Jurnal Teknologi Informasi dan Ilmu Komputer)*, Vol. 5.
- [5] Wang, Y., & Fan, W. (2018). A hybrid feature selection method based on naive Bayes and rough set. In *2018 International Conference on Big Data and Smart Computing (BigComp 2018)*, pp. 462–465.

-
- [6] Alqahtani, S., & Al-Somali, S. (2018). Text mining techniques: A review. In 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET 2018), pp. 1-6.E.
- [7] Mitra, P., & Liu, H. (2019). Sentiment analysis in social media: A review. In 2019 IEEE International Conference on Big Data (Big Data 2019), pp. 3204-3213.
- [8] Nurjanah, T. S., & Insanudin, E. (2016). Hack Database Website Menggunakan Python dan Sqlmap Pada Windows.
- [9] Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Tiktok. *Smatika Jurnal*, 10(02), 71-76.
- [10] Chen, D., & Kao, B. (2019). A hybrid feature selection method based on naive Bayes and ant colony optimization. In 2019 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2019), pp. 577-582.
- [11] Witten, I., Frank, E., & Hall, M. (2011). Data mining: Practical machine learning tools and techniques (3rd ed.). San Francisco: Morgan Kaufmann Publishers. Nurjanah, T. S., & Insanudin, E. (2016). Hack Database Website Menggunakan Python dan Sqlmap Pada Windows.
- [12] Firdaus, A. F., & Firdaus, W. I. (2021). Text Mining Dan Pola Algoritme Dalam Penyelesaian Masalah Informasi:(Sebuah Ulasan). *JUPITER (Jurnal Penelitian Ilmu dan Teknologi Komputer)*, 13(1), 66-78.
- [13] Rahman, N. A., Tan, K. L., & Lim, C. K. (2017). Predictive analysis and data mining among the employment of fresh graduate students in hei. *The 2nd International Conference on Applied Science and Technology 2017 (ICAST'17)*, 1-7.
- [14] Thomas, J., & Cook, D. (2018). Tiktok data mining: A handbook for journalists. New York: Routledge.